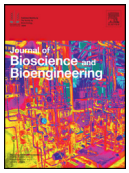

Journal of Bioscience and Bioengineering
VOL. 116 No. 3, 309–312, 2013

ELSEVIER

www.elsevier.com/locate/jbiosc

Genomic sequencing-based detection of large deletions in *Rhodococcus rhodochrous* strain B-276

Seikoh Saitoh,[1,†] Hiroaki Aoyama,[1,2,†,*] Masako Akutsu,[1] Kazuma Nakano,[1] Naoya Shinzato,[1] and Toru Matsui[1]

Tropical Biosphere Research Center, University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa 903-0213, Japan[1] *and Laboratory of Cell and Functional Biology, Faculty of Science, University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa 903-0213, Japan*[2]

Received 1 November 2012; accepted 2 March 2013
Available online 6 April 2013

Bacteria of the genus *Rhodococcus* (Actinomycetes) have the ability to catabolize various organic compounds and are therefore considered potential genetic resources for applications such as bioremediation. We investigated a next-generation sequencing-based procedure to rapidly identify candidate functional gene(s) from rhodococci on the basis of their frequent genome recombination. The *Rhodococcus rhodochrous* strain B-276 and its alkene monooxygenase (AMO) gene cluster were the focus of our investigation. Firstly, 2 types of cultures of the *R. rhodochrous* strain B-276 were prepared, one of which was supplied with propene, which requires AMO genes for its assimilation, whereas the other was supplied with glucose as the sole energy source. The latter culture was anticipated to have a lower gene frequency of AMO genes because of their deletion during cultivation. We then conducted whole genome shotgun sequencing of the genomic DNA extracted from both cultures. Next, all sequence data were pooled and assembled into contiguous sequences (contigs). Finally, the abundance of each contig was quantified in order to detect contigs that were highly biased between the 2 cultures. We identified contigs that were overrepresented by 2 orders of magnitude in the AMO-required culture and successfully identified an AMO gene cluster among these contigs. We propose this procedure as an efficient method for the rapid detection and sequencing of deleted region, which contributes to identification of functional genes in rhodococci.

[**Key words:** *Rhodococcus*; Pyrosequencer; 454 GS Junior; Deletion detection; Genome comparison]

Rhodococcus (Actinomycetes) are known for their ability to degrade a broad range of organic compounds, including short- and long-chain alkanes and aromatic compounds (1). Their catabolic activities have attracted interest for applications such as desulfurization of fossil fuel (2), enantioselective synthesis of compounds (3), and biodegradation of hazardous pollutants such as polychlorinated biphenyl (4), trichloroethene (5), and thiodiglycol, a chemical weapon hydrolyzate (6). These functions are often mediated by gene products encoded by large plasmids harbored in the cell (7–10).

Microorganisms readily undergo loss of genes not required for survival under specific environmental conditions (11). Under such conditions, mutant strains of *Rhodococcus* often arise, which are deficient in certain gene(s) due to loss of an entire plasmid or due to partial deletion of the encoding region. For example, *Rhodococcus rhodochrous* (formerly *Nocardia corallina*) strain B-276 can grow on propene as the sole carbon and energy source with alkene monooxygenase (AMO), an enzyme responsible for its assimilation (3,5). Saeki and coworkers (5) showed that AMO-deficient variants of this strain, obtained under non-selective condition, had undergone partial or complete loss of a 185-kb plasmid (named pNC30) encoding the AMO gene cluster. Identification of such plasmids and analysis of their sequences are therefore important steps for the development of useful *Rhodococcus* strains.

Recent years have seen the growth in popularity of the inexpensive next-generation sequencing (NGS) Roche 454 GS Junior system, which allows small laboratories to avail of high throughput sequencing technology. While its sequencing throughput is an order of magnitude smaller than that of higher grade models such as the 454 GS FLX system, the much smaller cost per run (~15% of the FLX) and its sufficient capacity for analyzing microbial genome (35 Mbp per run) make it a useful tool for microbial genomics. Using contig data from whole genome shotgun (WGS) sequencing performed using this system, we describe here a time-saving procedure for highlighting genomic differences between a wild-type strain and a variant deficient in gene(s) lost by deletion of a coding region in a plasmid.

Our procedure is based on the fact that, in a microbial culture growing under conditions in which a plasmid-encoded gene is required for growth, the population of cells harboring the plasmid will be overrepresented compared with another culture grown under conditions in which the gene is not required. This difference

* Corresponding author at: Tropical Biosphere Research Center, University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa 903-0213, Japan. Tel.: +81 98 895 8943; fax: +81 98 895 8944.
E-mail address: h085736@comb.u-ryukyu.ac.jp (H. Aoyama).
† The first two authors contributed equally to this work.

1389-1723/$ – see front matter
http://dx.doi.org/10.1016/j.jbiosc.2013.03.002

can be detected by comparison of WGS data between the 2 cultures with respect to the number of reads corresponding to the plasmid (Fig. 1). We sequenced WGS libraries prepared from 2 cultures of the *R. rhodochrous* strain B-276 whose sole energy sources were either propene, which requires AMO gene for its assimilation, or glucose [hereafter referred to as the selective condition (SC) or the non-selective condition (NSC), respectively]. Sequence data from the 2 WGS libraries were pooled and assembled to construct consensus sequences (contigs). The number of reads comprising each contig was counted, and the difference between the 2 libraries was analyzed statistically to detect overrepresented contigs. Finally, the overrepresented contigs were analyzed for the presence of the AMO gene cluster.

MATERIALS AND METHODS

Bacteria and growth conditions Cultures of *R. rhodochrous* B-276 (ATCC31338) were grown on Nutrient Broth no. 2 (Oxoid, Basingstoke, Hampshire, UK) containing 1% (w/v) glucose for the NSC, which was the condition used by Saeki and coworkers (5). For the SC, propene was supplied with propene (10%, v/v) via the gas phase with incubation in minimal medium (12). Cultures were grown in shaking flasks at 30°C with aerobic shaking (180 rpm) under dark condition for a week. Cells were harvested by centrifugation (8000 ×*g* for 10 min), washed twice with sterilized water, followed by re-centrifugation (8000 ×*g* for 10 min) to obtain cell pellets for DNA extraction. Cell pellets were stored at −80°C until use.

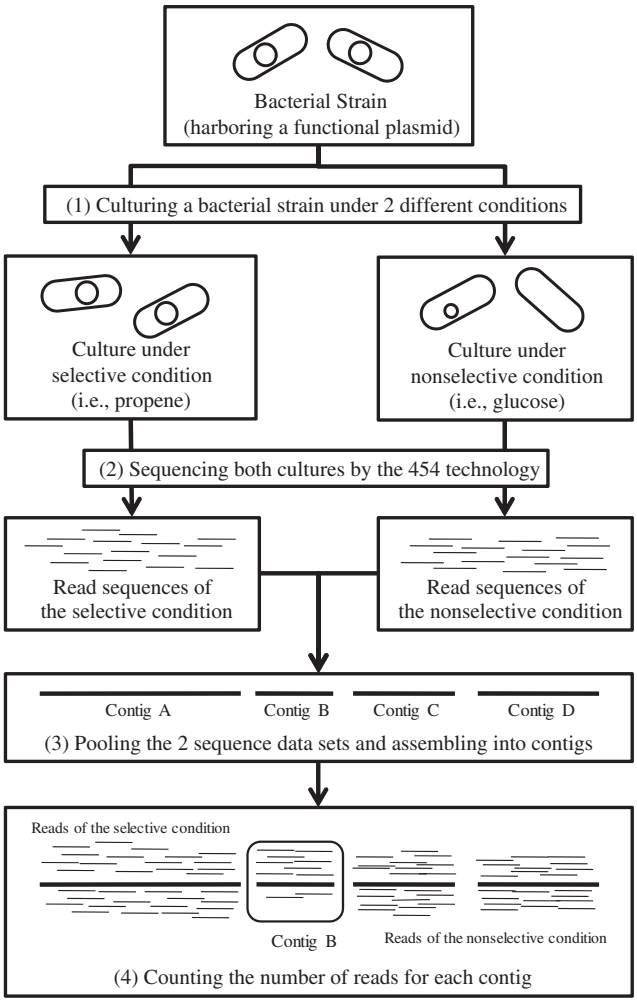

FIG. 1. Conceptual scheme of the procedure to detect large deletions in a bacterial genome, as suggested in this study: 1, A bacterial strain is cultured under 2 different conditions (feeding propene or glucose as the energy and carbon sources); 2, genome sequence from both cultures is analyzed by whole genome shotgun sequencing by using the 454 pyrosequence technology; 3, the 2 sequence datasets are pooled and assembled to construct a set of contigs; 4, the number of reads for each contig is counted, and the contig(s) whose abundance of read sequences is highly biased between the 2 libraries would be detected.

TABLE 1. Estimated proportions of cells capable to grow on propene of cultures after incubation under SC and NSC.

Culture	Number of colonies	Grown	Proportion (%)	
			Estimated	95% CI[a]
NSC	20	0	0	0–17
SC	22	22	100	85–100

Summary of growth test in which propene was supplied as the sole energy source was shown (see text for the procedure).
[a] 95% confidential interval based on exact confidential interval (13).

Examination of proportion of cells capable to utilize propene in the cultures after incubation Culture after incubation under SC or NSC was examined in terms of proportion of cells capable to grow on propene as the sole energy source. The test was conducted using an air-tight glass vial with 3 mL of the minimal medium, 28.5 mL air, an inoculum from a colony of SC or NSC, and 20 mL of propene injected using a syringe. Each vial was incubated at 30°C with aerobic shaking. The criterion for the test was whether OD_{600} measured after incubation for 96 h exceeded 0.8 or not. A 95% confidential interval of the proportion was calculated using exact confidential interval (13).

DNA extraction and shotgun pyrosequencing Bacterial genomic DNA was extracted using a DNA extraction kit, ISOPLANT I (Nippon Gene, Tokyo, Japan). After extraction, the DNA was quantitatively and qualitatively analyzed by a Nanodrop (Thermo Fisher Scientific, Wilmington, DE, USA) and Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). NSC and SC DNA samples were subjected to shotgun pyrosequencing by using a 454 GS Junior titanium chemistry (454 Life Sciences-a Roche Company, Branford, CT, USA), and DNA libraries were thus prepared. Pyrosequencing runs were carried out in duplicate for the NSC library and once for SC library. All operations were carried out according to the protocols provided by the manufacturers.

In silico analyzes Sequence data were pooled and assembled using Roche Newbler software (ver. 2.6) (14) with default settings. The abundance of DNA corresponding to the contig sequence in the 2 libraries was derived from the number of reads. For standardization purposes, the number of reads per 1 kbp sequence per 1 M aligned reads (hereafter referred to as RPKM) was used to compare the abundance of sequences between the 2 libraries. This analysis was performed using an in-house script in the R statistical software (ver. 2.13, R Core Development team). The difference of abundance between the 2 libraries for each contig was examined using the Fisher exact test based on the total aligned number of reads. *P*-values were adjusted with Bonferroni correction.

The Basic Local Alignment Search Tool (BLAST; ver. 2.2.25) (15) was used to identify contigs containing the AMO gene cluster (GenBank: D37875.1) (16). The bb.454contignet Perl script (17) was then used to illustrate paths between the contigs assembled by the Newbler software.

A putative annotation of the constructed draft genome sequence was made using Microbial Genome Annotation Pipeline (MiGAP, an auto annotation pipeline of DDBJ, http://www.migap.org/) for discussion on similarity of the genome to other genomes of rhodococci sequenced ever.

The sequence data from the present study have been submitted to the DDBJ Sequence Read Archive (DRA) under accession no. DRA000810.

RESULTS AND DISCUSSION

In the present study, we cultured *R. rhodochrous* strain B-276 grown under 2 different conditions (NSC and SC). Result of growth

TABLE 2. Statistic of sequencing runs using 454 GS Junior.

	Sequencing run no.		
	1	2	3
Library[a]	NSC	NSC	SC
Sequencing throughput			
Total number of reads	127157	126816	121990
Total number of bases	57886794	59169882	55981563
Assembly result[b]			
Number of aligned reads	126048	125660	121167
Number of aligned bases	57484861	58743091	55750732
% Aligned base	99.31	99.28	99.59

[a] The non-selective condition (NSC) library or the selective condition (SC) library.
[b] Data from all 3 runs were pooled and assembled by Roche Newbler (ver. 2.6) using default settings.